Ravi Prakash

Professional with several years of full time experience in Backend Software Development, MLOps, IOT across several product-based startups. Ability to hack around multiple technologies and build highly scalable, low latency, concurrent, distributed systems, RESTful services

9868576600

prakashravip1@gmail.com
Github
LinkedIn
Blog

EXPERIENCE

BatterySmart, Gurugram — *SDE-3*
Jan 2023 - Present

- Develop EV battery's IOT infrastructure by collecting IOT events from multiple battery vendors and storing them in a Apache Pinot realtime table. It is able to handle scale of over 70 qps and manage TBs of data storage and power real time analytics. Tech used – Golang, Kafka, Avro, Apache Pinot, Prometheus, Grafana, Redis, Trino, Redash, Kubernetes
- Written a kafka producer to send Avro serialized electric battery IOT event data for a vendor. Tech used - Python, Avro, Kafka, Kubernetes
- Developed an Insights service which captures IOT event for every EV Battery swap transaction and stores it in a Apache Pinot realtime table. Tech used – Golang, Kafka, Pinot, Kubernetes
- Developed showing live EV battery location based on IOT data. Tech used: NodeJS, Redis, Kubernetes, Mysql, Redis Geolocation
- Developed modules for B2C order delivery service -
  - module for storing bulk driver location pings to used by driver availability.
  - module for calculating price for order delivery based on multiple parameters.
    Tech Used: Golang, go-gin, gorm, Postgres Kubernetes

magicpin, Gurugram — *Tech Lead*
Sept 2022 - Jan 2023

- Took lead in setting up org level observability using Prometheus and Grafana.
  - Deployed in house Grafana OSS in Google Kubernetes Cluster.
  - Deployed prometheus server in GKE cluster for scraping prometheus exporters from GCE autoscale VM instances and GKE cluster deployed pods, leveraging prometheus service discovery
  - Helping software teams with prometheus client integration into their Go, Java, Python applications.
  - Reducing organisation wise metric samples ingestion by applying metric relabeling filters.
- Designing, developing a backend system for serving merchant ad package pricing logic.
- Designing, developing a real time ML prediction and serving architecture for Google Vertex AI trained statistical ML model. Design
  Technology used: fastapi, pydantic, aiokafka, bigquery, bigtable, gcs
  - Building model prediction and feature store in Google nosql storage – bigtable.
  - Created ORM over pydantic data models to facilitate bigtable column family read and write.
  - Setting up a pipeline to do model feature calculation and storage to bigtable feature store.
  - Setting up a pipeline to do model prediction on several triggers.
  - Setting up data sync pipeline between bigquery tables to bigtable for feature calculation
  - Setting up kubenetes schedule jobs to do model's score percentile calculation and feature and data ingestion to Google bigquery for analytics.
  - Setting up API servers that do real-time model prediction for the end user.
  - Creating APIs to serve predictions done from multiple sources.
  - Integrating open telemetry in API middleware to collect request span and

TECHNICAL SKILLS

Golang, Python, C/C++

Delta Lake

Cassandra, Bigtable, DynamoDB, Apache Pinot

MongoDB, ElasticSearch,

Kafka

MySQL, Postgres

Redis, Aerospike

Go-gin, go-mux, gorm, Pandas, Django, Flask, FastAPI, Sanic, Keras, Computer Vision, Asyncio

AWS, Azure, GCP

Docker, Kubernetes

Prometheus, Datadog, Grafana

AWARDS

Gold Medal - Secured 1st State Rank and 20th International Rank in International Informatics Olympiad (IIO) in 2010.

CBSE Certificate of Merit: Secured grade A1 in all 5 subjects in Xth Board Exam.

exporting to Google Cloud Trace.
    - Setting up bitbucket CI/CD workflows for deployment to kubernetes.

Tekion Corp, Bengaluru — *ML Engineer*

July 2021 - Sept 2022

- Driving designing and development of distributed batch and streaming ETL processing pipeline, offline and online ML model serving architecture.
- Responsible for setting up a scalable, cheap, ACID compliant delta lake storage over S3 for storing processed raw facts known as features and batch prediction results of ML models, replacing postgres. Blog
- Responsible for development and maintenance of several internal python clients and tools packages used by ML Operations and data science team.
- Instrumental in bringing modern software development principles in all ML data pipeline batch jobs and ML model inference APIs.
- Optimized top k recommendation API latency by doing batch cosine similarity calculation and serializing high dimension feature vectors in parquet and storing it in an in-memory based fast storage.
- Created a bare SQL parser over pyarrow for ML data pipeline to facilitate direct read of raw data from delta lake warehouse. It supports query filtering over partitioned and non-partitioned columns, thereby reducing the need and cost of querying a managed query engine.
- Developed a scalable architecture for sharing duplicate data merge pipeline results using spark, delta lake, Kafka, and AWS Athena.
- Splitting sales chat app built using RASA framework into containerized (core and ACTION) microservice to facilitate communication among them using custom API routes and integrated our observability tools and dynamic model loading.
- POC for automating the manual task of vehicle data induction in Tekion from an OEM website using selenium, beautiful soup and Kafka.
- Added concurrency to ML serving APIs by converting all I/O bound tasks into asynchronous calls and using ASGI supporting web frameworks like FastAPI and parallelism by multiple uvicorn workers integrated with open-source elastic stack observability. [opentelementry, elastic-apm, ElasticSearch, Kibana]

daloopa, Noida — *Software Engineer III*

Dec 2020 - May 2021

Technology used: Django, GraphQL, AWS ECS Fargate, Celery, Azure Container Apps, MySQL

- Researched and implemented foreign language OCR detection, recognition and translation of financial documents using open source technologies thereby saving costs.
- Designed and implemented a data pipeline to support pattern-based search on financial documents using machine learning, OCR, NLP and heuristics-based search algorithms.
- Automating manual tasks through AWS batch jobs.
- Improving latency of critical flows through caching, concurrency, multiprocessing and pre-processing.
- Configuring dockerized ECS Fargate cluster pipeline for development and production

magicpin, Gurugram — *Software Engineer*

June 2018 - Nov 2020

- Responsible for continuously evaluating recent technologies and driving architecture and design decisions on various critical backend systems and ownership to run and maintain cloud infrastructure.
- Took initiative for dataset collection, feature engineering, training using transfer learning, model inference using RESTful flask service on CNN based deep learning classification model for selfie classification.

Central Sector Scheme for College and University Student
Received scholarship for securing high percentile in JEE Main

Courses

Introduction to Machine Learning in Production

Specializations

Programming with Google Go by University of California, Irvine

Google Cloud Platform Fundamentals: Core Infrastructure

DeepLearning.AI TensorFlow Developer

- Migrated 21 TB data from 21 node Cassandra cluster to Cloud Bigtable cluster with redesigned schema ensuring row key entropy
- Migrated a tightly coupled java-MySQL monolith backend stack to decoupled, low latency, go service with denormalized MySQL schema thereby improving availability and reducing infra costs.
- Designed and developed a new low latency, highly available reverse-proxy layer between API layer and front-end for several backend systems which resulted in secure and robust backend systems.
- Designed and developed concurrent 5 layer trending pipeline which generates scores accumulated over days with decaying functions for various social entity in our app like posts, hashtags, merchants, users, friends and using worker-pool, pipeline, barrier design patterns
- Designed and Implemented Consumer and API Flow for projects like Merchant Rating calculation based on Bayes estimation , merchant complaints and review flow and Hashtag Feed and Campaign Feed and selfie and hashtag approval flows, groupbuys, merchant and user feed
- Designed and developed notification,sms,whatsapp,IVR trigger logic along with an order cleanup service for Order Management Team

Constems AI Systems, IIM Lucknow — *Software Engineer Intern*

May 2017 - July 2017, Nov 2017 - Dec 2017

Worked as part of the Product Development Team of 3 and made valuable contributions to the following products by developing web and android apps

- RetailVision – Gender Recognition, Real-Time Tracking of Customers
- StaffLogin – Face Recognition based Staff Login System

EDUCATION

NSIT, New Delhi — *Bachelor in Engineering*

2014 - 2018

Branch - Computer Engineering - CGPA 8

BGS International Public School, New Delhi — *Class 6 to 12*

2007 - 2014

12th CBSE Board - 93.8 %

10th CBSE Board - CGPA 10